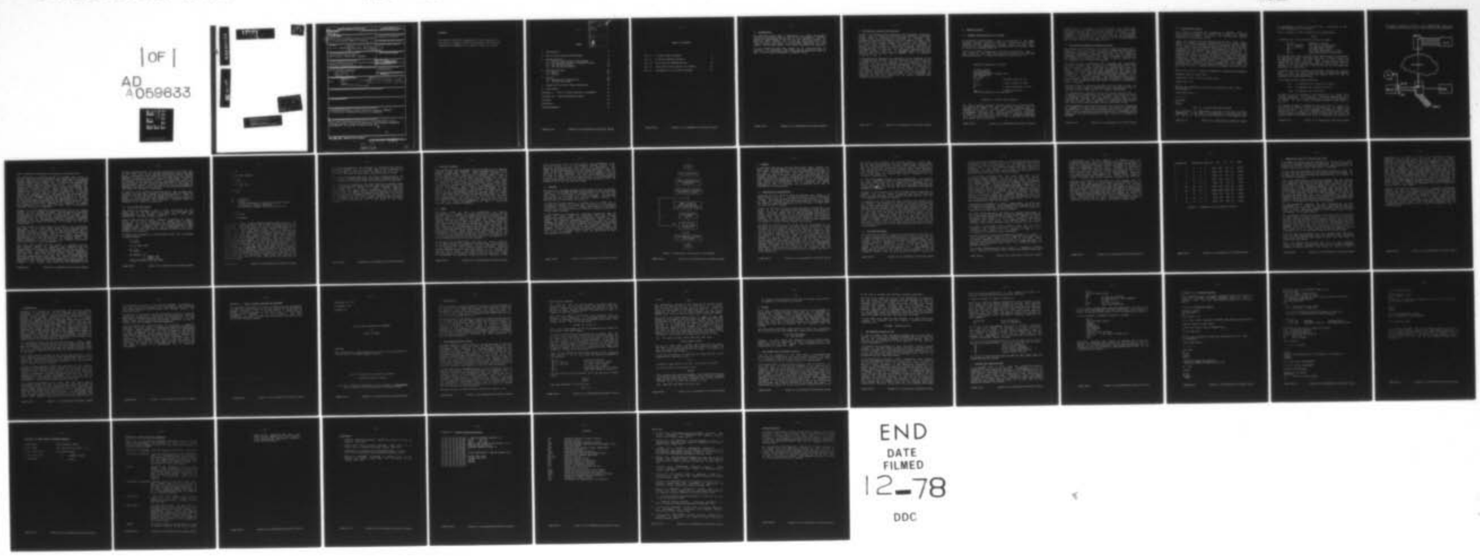
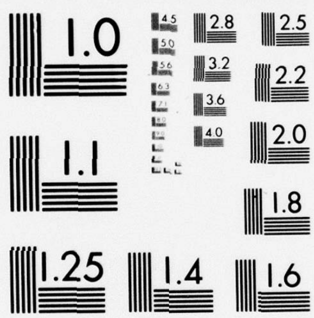
AD-A059 633 UNIVERSITY COLL LONDON (ENGLAND) DEPT OF STATISTICS --ETC F/G 9/2
ACCESS TO AN INFORMATION RETRIEVAL SYSTEM VIA A DISTRIBUTED COM--ETC(U)
DEC 77 A V STOKES
UNCLASSIFIED TR-40 NL

| OF |
AD
A059633



END
DATE
FILMED
12-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (14) TR-40	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Access to an Information Retrieval System via a Distributed Computer Network,		5. TYPE OF REPORT & PERIOD COVERED Technical 1975-1977
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dr. A.V. Stokes (10) Adrian V. Stokes		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS University College London, Gower Street, London WC1, ENGLAND		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Defence Advanced Research Projects Agency 1400 Wilson Boulevard, Arlington, Virginia 22209, USA		12. REPORT DATE (11) 31 December, 1977
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research (12) 48p.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE U

16. DISTRIBUTION STATEMENT (of this Report)

No.1

This document has been approved for public release and sale; its distribution is unlimited.

(9) Technical repts

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Network Measurements, Packet Switched Networks, ARPANET, Bibliographic Information Retrieval and MEDLINE.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes an analysis of the access made to the National Library of Medicine databases, particularly CANCERLINE and MEDLINE, by a group of users in the UK.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

78 09 11 049

409 714

slt

ABSTRACT

This Report describes an analysis of the access made to the National Library of Medicine databases, particularly CANCERLINE and MEDLINE, by a group of users in the UK.

INDEX

INDEX TC DIAGRAMS

1. Introduction

In a previous report (Ref. 1) describing the INDRA measurement project, we mentioned various aspects of the programme. One such aspect was measurement of the use of the National Library of Medicine (NLM) system; due to data not being available at the time of writing that report, it was briefly described but only sample results were given. This report rectifies that omission.

As it is preferable that this report be as self-contained as possible, it reiterates some descriptions previously given in Ref. 1, particularly regarding the system itself.

2. The British Library STEIN Project

In 1973, the British Library Research and Development Department (BLRD), then the Office for Scientific and Technical Information (OSTI), began a comprehensive programme of research into information retrieval by means of on-line computer systems. This project consists of three phases and it is with two aspects of the short term project (STEIN) with which we are concerned in this report (Refs. 2 and 3). One consisted of an appraisal of several on-line systems. Of these, one - MEDLINE - was accessed via ARPANET and, in addition to providing documentation and advisory services (Ref. 4), we provided access and evaluated usage in addition to other usage of ARPANET. This is described in detail in Ref. 1.

The second aspect concerned an evaluation of two databases at NLM - CANCERLINE and MEDLINE - and the background to this project is described in Ref. 5. In addition to the facilities already mentioned, we developed extensive measurement tools and, by use of these, evaluated the use of the NLM system. In particular, we conducted a week-long experiment which enabled us to collect a reasonable body of data and it is this that forms the basis of this report although other available data are included. In addition, the usage of the dial-up lines to the LONDON-TIP was monitored by the Post Office and we mention relevant data obtained from that monitoring.

3. System Overview

3.1 General Description of UCL System

At University College London, a PDP-9 is connected to the ARPA Interface Message Processor (IMP) via a Very Distant Host (VDH) interface and the measurement programs reside in this Host. There are two major measurement programs, QUES and MEDLINE. Both are written in Babbage (Ref. 6).

QUES is intended to interrogate users of the London-TIP. When a user dials up the TIP, QUES intercepts that user and asks him various questions. A typical scenario is shown in Fig. 1.

```
LONDON-TIP MONITORING SERVICE.

SURNAME >stokes
TIP PASSWORD >xxx
PASSWORD UNKNOWN - REENTER >ucl
HOST NUMBER >42
OK - BYE
Closed                            * message from the TIP

@O 42                             * user connects to host
.......
@C                                * close connection to host

Open                              * message from TIP
NEXT HOST NUMBER OR RING OFF NOW >
```

Figure 1: A Typical QUES Scenario

The data recorded by QUES, together with various control data, are punched onto paper tape (thus obviating problems such as closing files after a system crash). This data includes the information given in the initial interrogation (suitably datestamped) together with information when the user logs off. In addition, every half hour, QUES prints a message indicating that it is still active; this is to help in determining the time of a crash, should one occur.

In the case of a connection to a standard Host, the measurement program takes no further action. However, connections to NLM raise many problems and a program was written to simplify this procedure. Furthermore, it was required to measure various parameters in depth and hence a greater degree of interaction with the user was required. For this reason, the MEDLINE program was written and this is described below.

3.2 The National Library of Medicine System

The hardware at NLM consists of two IBM 370/158 processors with 4 Mbytes of core and a very considerable amount of disk storage (5600 Mbytes, mainly 3330s). The major use of the system is providing information retrieval facilities and of collating information from the medical literature. One major activity is the production of "Index Medicus". This is typeset directly from the computer using a Photon 901 machine.

The data for such a publication are available for information retrieval systems and is known as the MEDLARS data base. This may be accessed in either batch mode or interactively (and, in fact, a batch service has been available in the UK for some years). The batch service has been available at NLM since the beginning of 1964 with the interactive service available since October 1971. The number of searches performed annually is of the order of half a million. The interactive system uses a program ELHILL3 derived from the System Development Corporation (SDC) ORBIT system and the combination of ELHILL3 and MEDLARS is known as MEDLINE.

Although MEDLARS is the major database on NLM, other systems are available, for example, CANCERLINE (sic), CANCERPROJ (a list of research projects into cancer) and TOXLINE (toxicology).

The communications processor for the NLM system (an IBM 3705) is connected via leased lines to five ports on the National Bureau of Standards (NBS) TIP. Thus the ARPANET "sees" NLM as five teletype-like devices and conversely, NLM sees ARPANET as five 2741-like devices. This interface is completely non-standard and, due to this, many problems arise as described below.

These problems are due entirely to the ad hoc way the connection to ARPANET was made and resulted from complex non-technical considerations. The problems described below do not occur with other hosts connected in a more orthodox way to ARPANET; nor do they occur with NLM connections to other networks e.g. TYMNET and TELENET.

3.3 The MEDLINE Program

In the case of a standard Host connection to ARPANET, there is very little likliehood of errors and various flow control mechanisms are provided. Because of the non-standard NBS/NLM connection, many problems arise.

First, the standard TIP flow control mechanism for human users, namely of sending control-G (bell) when it can no longer accept input, is not only of little use with an automaton (in that there is no user to hear an audible signal) but is, in fact, counter-productive in that the parity of the character is illegal to the communications controller and therefore causes I/O errors. In addition, the 2741 is a half-duplex device; ARPANET is, in general, full-duplex and again the receipt of data at wrong times leads to I/O errors. Finally, although in the case of a standard connection, there is a protocol (the Initial Connection Protocol - ICP) defined which allows users to connect to a standard "socket" then transfers them to an appropriate free socket, no such mechanism exists in the case of NLM and the user must try each port in turn. A typical series of commands is shown in Fig. 2.

```
WHEN I LAST LOOKED A COUPLE OF MINUTES AGO, MEDLINE WAS AVAILABLE

TERMINAL (FAST OR SLOW) >slow

PLEASE WAIT WHILE I TRY TO CONNECT YOU...

TRYING PORT #134

ID >
YOU ARE NOW CONNECTED TO PORT #134 AT MEDLINE AS USER  NLIC2
NLM PASSWORD >

HELLO FROM ELHILL3.

.............

GOOD-BYE!
@c
CLOSED
```

Fig. 2: A Typical MEDLINE Scenario

This procedure is complicated even more by the lack of status information. In the case of a standard Host connection, the IMP can report on the status of the connection - for example "Host

Not Responding", "Open", "Net Trouble" etc. In the case of NLM, no such information is available.

Thus, to connect to NLM, instead of the conventional:

@C nn

a sequence such as the following has to be typed:

@H 147	Connect to NBS TIP
@R F S 134C0003	Set receive socket number
@S T S 134C0C02	Set send socket number
@E I	Set mode (echo local)
@F	Flush any characters in buffer
@P B	Open both halves of connection

The TIP may then reply "Closed" which implies that the connection is currently in use. Sometimes, the reply is "Open" followed by the NLM system prompt. In this case, it is clear that the connection is open correctly. More frequently, the reply is "Open" followed by no message from NLM or, even worse, no TIP message. In this case, it is tcally unclear what is happening, especially to naive users.

To obviate all these problems, the MEDLINE program was written (Refs. 7 and 8). The program has been considerably modified since its first specification by the author (Ref. 7) and we describe the latest version only.

The program performs three functions:

(i) it simplifies the connection process

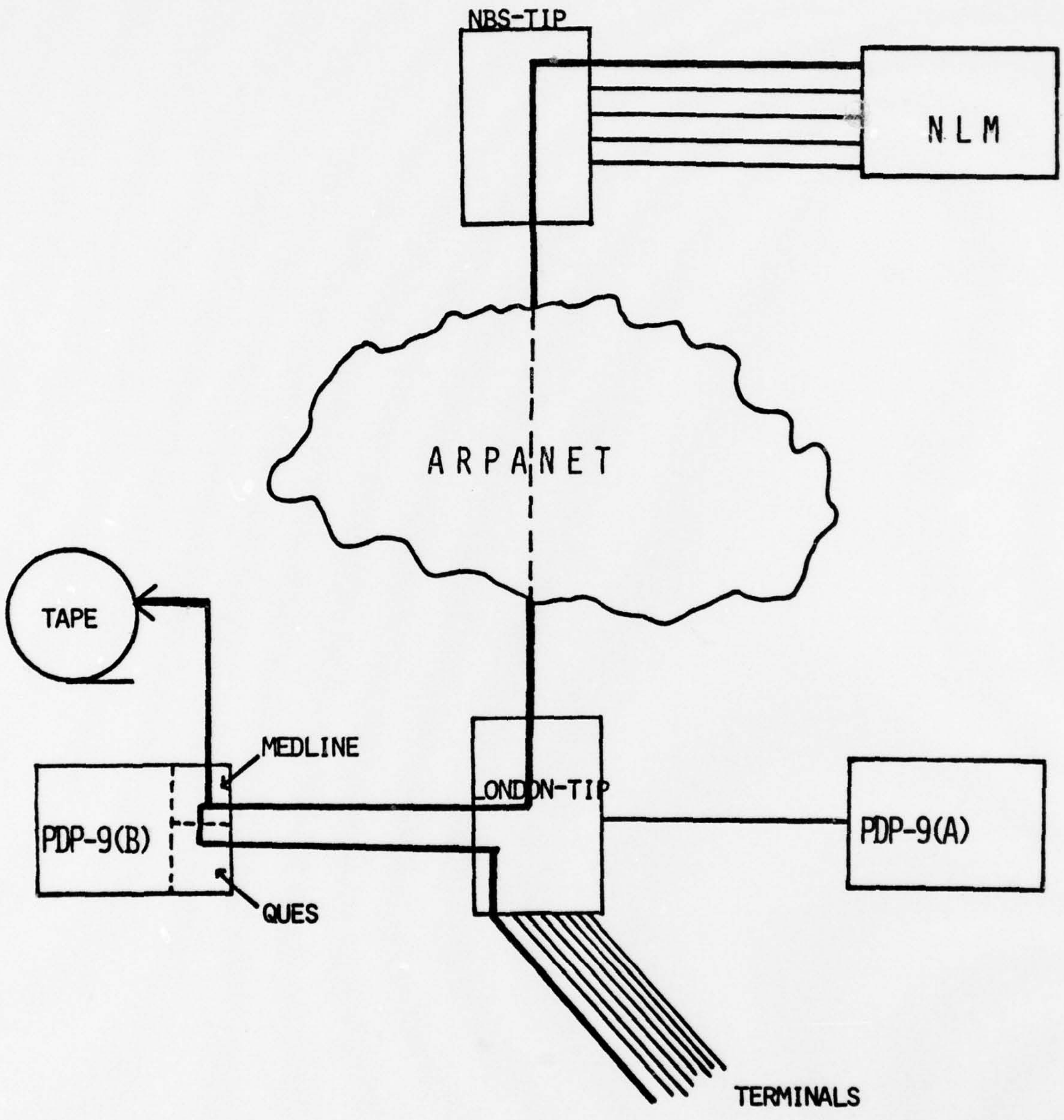
(ii) it modifies the dialogue with NLM

(iii) it monitors the interaction

In order to use the program, a user connects to the INDRA PDP-9 (either directly, in the case of a non-dialup connection, or by specifying MEDLINE - or, equivalently, 147 (the NBS-TIP number) - to QUES). He then enters a dialogue similar to that shown in Fig. 2.

The program attempts to connect to the five ports in turn, but the order in which it tries is determined by the speed of the user's terminal. It will, however, attempt to make the connection to a port of a different speed rather than to none. If it is unable to connect, it tries the next port and so on. By having access to the Host-Host protocol responses, the program is

FIGURE 3: DATA PATH FOR MEDLINE ACCESS



able to determine accurately the status of the connection.

If a connection is correctly opened, the program then tries to log in. This may not be trivial in that, again due to the nature of the connection, a previous user may have been disconnected without having been logged out. Therefore, the program must "tidy up" the connection before logging in. It achieves this fairly complex task by operating as a finite state machine with, currently, seventeen states. After sending the LOGIN, NLM prompts for an ID and this is relayed to the user. If the user specifies one, this is sent; if not, signified by a carriage return, the program selects one from the five available to UK users (more specifically, it only selects one which it believes is not in use). If NLM indicates that the chosen ID is, in fact, in use (for example, by a user who has preferred not to use the UCL program), the program marks it so and chooses another. When it logs in correctly and is prompted for a password, it passes this request on to the user and from then on maintains a virtually transparent connection.

However, it is not completely transparent and this is the second feature of the MEDLINE program. The NLM system is extremely verbose (even in experienced user format) and the program tries to minimize this to some extent. In particular, the PROG: prompt which indicates the start of text from NLM is deleted completely (together with the accompanying carriage return/line feed) and the USER:<cr><lf> prompt to the user is replaced by the more conventional >.

The final function of the program is to monitor the interaction with NLM. It is clear that, since all the dialogue passes tthrough the PDP-9, it is possible to copy this dialogue onto a suitable recording medium. This is done, each line having the date and time (to the nearest tenth of a second), the channel designation (A to K but since A is the QUES control channel, this should never occur) and the direction of transfer added. The system schematic is shown in Fig. 3 and a sample of the output from the program is given in Appendix II.

There were some problems over the choice of recording medium. The obvious choice of magnetic tape suffered from the disadvantage that it was not always possible to tidy up after a system crash. Although the system dump routine was modified to close the file, the routine was not always called. Therefore, the file was often not closed and this presented the tape software re-opening the tape when the system auto-restarted. This problem was solved by modifying the open routine so that, on detecting a file which had not been closed (more specifically, on finding a file trailer label which did not correspond to the

preceding header label), the routine treated the last valid file as the last file on the tape and proceeded to write from that point. Occasionally, this algorithm also failed and was further modified so that, in the event of a tape open error, the program changes the monitoring medium to drum. In this event, there is no problem in the case of a crash. The reason that drum is not used in general for monitoring is that the size of the drum is significantly smaller than the tape and it is not convenient to transfer from drum to the RL 360/195 at the Rutherford Laboratory (RL) where the analysis is performed.

In addition to the functions described above, the program also provides the user with a number of useful facilities. These are access to the date and time, the system status (i.e. a list of logged in users) and the time logged in in the current session. There are also facilities for providing a message which is broadcast to all users on logging in.

3.4 Use of the NLM System

This section is intended to give a brief description of the actual use of the NLM system in order to understand the significance of the results presented later. For more information, reference is made to Ref. 9.

Lines to NLM consist of either commands (indicated by double-quotes as the first character e.g. "PRINT FULL) or search statements. In the case of MEDLINE, the sub-terms in the search statements must be selected from a controlled vocabulary - MeSH (Medical Subject Headings, Ref. 10). CANCERLINE is a free language data base and may be searched on words in the title, abstract or on keywords.

To explain the elements of a session more easily, let us consider a simple example:

```
SS 1 /C?
>  rabies

SS (1) PSTG (397)

SS 2 /C?
>  france

MM (FRANCE) (2)
            1  FRANCE (MH)
            2  FRANCE (PI)
SPECIFY NUMBERS, ALL, OR, NONE-
```

```
> all

SS (2) PSTG (26330)

SS 3 /C?
> 1 anD 2

SS (3) PSTG (22)

SS 4 /C?
> "prt

1
AU  - Mueller WW
AU  - Schoop U
TI  - Natural resistance of an African rodent Pracmys
      natalensis to rabies infection.
SO  - Ann Microbiol (Paris) 127(3):447-53, Apr 76.

SS 4 /C?
>  wertyujkl

NP (WERTYUJKL)
```

This brief extract from a terminal session shows enough of the features of a search to enable the results presented below to be understood. The first line, "SS 1 /C?" is a prompt from NLM to ask the user to input either his first search statement or a command (in the "new user" format, this is spelled out explicitly, at very obviously a considerable increase in connect time). The user then input a simple term (a MeSH heading) "rabies" and the program replied that there were 397 citations ("postings") with "rabies" as a keyword. NLM then prompted for the second search statement. On being given the word "france", it indicated that this had more than one meaning ("MM" means "Multi-meaning Message") and the user was asked to specify which was required. On stating that it was required both as a MeSH heading (MH) and a place name (PI), NLM found 26330 citations. The user's search statement 3 indicated that he required the set of all citations in those given by (1) and (2). He could have given this in response to the first prompt by specifying "rabies and france", but it is frequently found that users prefer to input sub-terms one at a time. This pattern of usage will be discussed later.

The program responded to the statement by indicating that only 22 citations existed with both terms and, on being prompted for search statement 4, the user asked the system to print the 22 citations. In order to save space, only the first is given.

After the printing has been done, the user is again prompted for SS4 (since a command was given rather than a search statement). As the input is not a valid term, the system finds "No Postings".

This search is not typical of the vast majority of searches which are considerably more complex and may take very much longer time (the average for a search is about fifteen minutes). Furthermore, the actual search statements may be made more complex, using the "AND", "OR" and "NOT" connectives more freely. Finally, there is a command "DIAGRAM" which asks NLM to print out the current state of a search as a tree diagram and the users were requested to use this command before the print command. This was done in many cases, but is not shown in the above example.

4. Analysis Programs

Two programs are used to analyze the data produced by the MEDLINE programs, DEMUX and MEDLINE. These programs form part of the XPSTATS package (Refs. 1 and 11); this package consists of a driver routine which performs many functions common to many of the segments (e.g. it checks that the date/time tags are in order, irrespective of the three possible formats). In brief, its operation is as follows. It initializes itself then reads its tables; it then reads and parses (Ref. 12) a control file specifying the operations required, their parameters and other control information. This control file effectively selects a function - for example, DEMUX. The appropriate segment is then called at an initialization entry point. The appropriate data file is selected and scanned until the initial date is found. Until the final date is found, the appropriate segment is called at its main entry point for each line. On finding the end date (or, equivalently, the end of file or, in rare cases, an irrecoverable I/O error), the segment is called once more to tidy up. This is shown diagrammatically in Fig. 4. Let us now describe these programs in detail.

4.1 DEMUX

The first, DEMUX, is used to demultiplex the various conversations. A number of interactions may be occurring at any one time (in fact, due to the number of internal channels in the PDP-9, there is a limit of four people who can be interacting with NLM through the PDP-9 at one time) and each interaction consists of the user to PDP-9 and PDP-9 to NLM halves (which are very similar after the login is completed). Also, there are the data sent to the system console and the paper tape punch. All these interactions are datestamped and written to the tape (or other recording medium) in chronological order, with the channel designation added. DEMUX takes this multiplexed data and demultiplexes it into the appropriate streams. Since not all may be required, particularly as some may be virtually identical, the user may specify which of the four he requires - user/PDP-9, PDP-9/NLM, PDP-9/system console or PDP-9/punch

Furthermore, the program splits the data on various streams, where appropriate, into separate conversations. A conversation is defined as would be expected e.g. on the PDP-9/NLM channels, it is defined as that between /LOGIN and LOGOFF. Since the channel designation is by now superfluous, it is removed and replaced by a colon. Also, line reconstruction is performed. When the program sends data to tape, it ensures that the record is terminated by carriage return/line-feed. However, in some

cases, lines are sent to the user without such terminators (e.g. the ">" prompt) and so the MEDLINE program appends a "$" character to indicate that it has added the terminators. The DEMUX program reconstructs such part lines correctly so that the output is an exact transcript of the terminal record with the addition of the datestamp. The output is preceded by a row of "+" signs to enable the user (and the later MEDLINE analysis program) to detect the start more easily. It is terminated by various statistics such as the time for the interaction and the number of lines sent in each direction.

4.2 MEDLINE

The output of the DEMUX program is then taken as input to another segment of the XFSTATS package called MEDLINE (it is appreciated that there is a possibility of confusion over the use of this name since it now refers to three programs - the analysis program, the PDP-9 program and the NLM system; however, the meaning will be clear from the context).

The MEDLINE program operates in a similar way to DEMUX and it performs its analyses on however many conversations with which it is supplied. Usually this is a simple concatenation of the DEMUX output but this may be edited as required to select interactions of particular interest.

It then prints out the number of sessions encountered and the number of these where an interaction with NLM occurred (designated "non-trivial"). The sessions are analyzed in some detail and eighteen parameters evaluated. These are, for example, the number of search statements per session, the number of postings found per search statement and the time per line from NLM. In each case the total, mean and standard deviation are printed (although these may not be meaningful in all cases). In addition, a list of search words for which no postings were found and a list of commands used (with an indication of their occurrence) are printed.

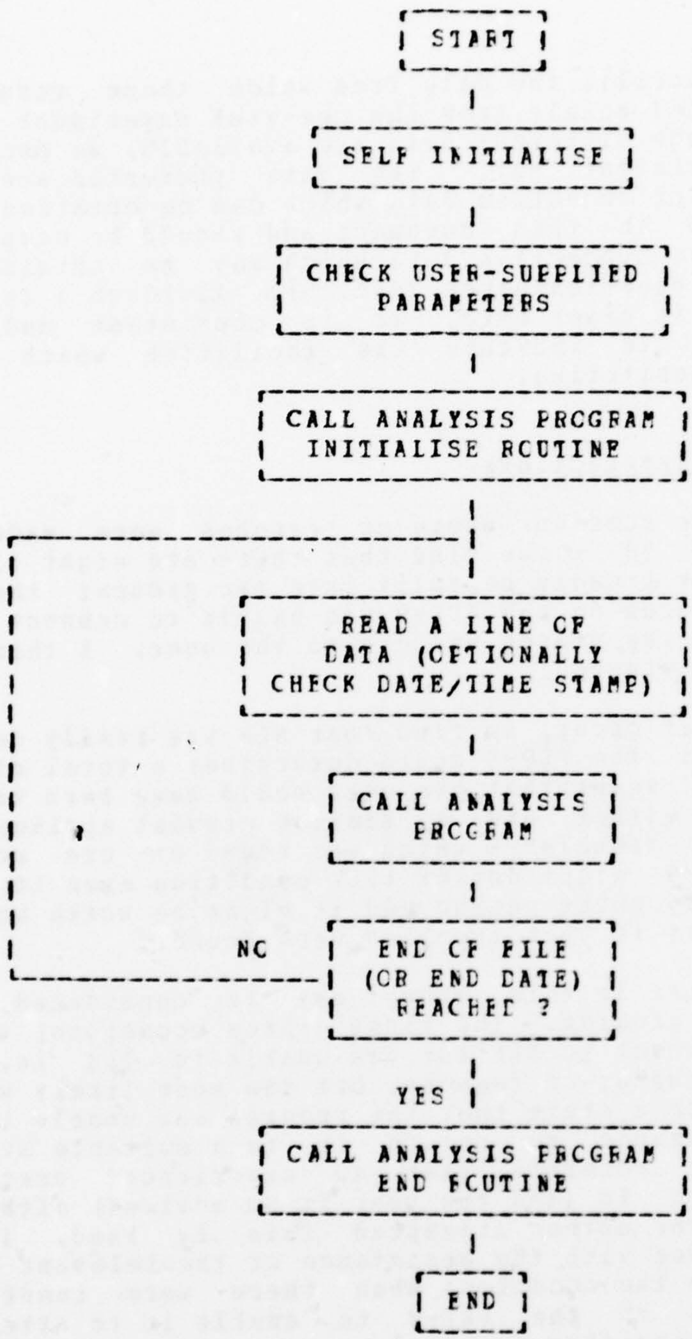


Figure 4: Generalized Flow Diagram for XPSTATS

5. Results

As explained previously, the data from which these results are drawn were obtained mainly from the one-week experiment early in 1977 since, although different data are available, we decided to use only a consistent set. The data presented are used to indicate the type of objective data which can be obtained by the methods described in this document and should be compared and contrasted with the subjective data which may be obtained from user reports and questionnaires (Ref. 5). Although a relatively small set of data is given here, it is consistent and serves sufficiently well to indicate the facilities which may be provided by this monitoring.

5.1 Unsuccessful Interactions

Examining first the sessions where no searches were made - of which there were 38 - we find that there are eight different reasons. These may broadly be split into two groups; the first is when the program on the PDP-9 was unable to connect to NLM; the second is when the reason was due to the user. A third cause was when the PDP-9 crashed.

Looking at the first group, we find that NLM was really down (at least, as far as the PDP-9 could determine) a total of twelve times. This cause meant that the user would have been unable to connect directly either and a similar comment applies to the second cause - "net trouble" - which was found on one occasion. However, the PDP-9 might detect this condition even if the net were down for a very short period and it might be worth modifying the program to retry if "net trouble" were found.

The other two reasons in this group may be considered to be failures of the program. The first (three occasions) was when the PDP-9 could connect to NLM but was unable to log in. This could be due to a number of reasons, but the most likely was that the port was in such a state that the program was unable to send the requisite commands to return it to a suitable state for logging in. It is possible that an experienced user could succeed in logging in (and the user is so advised) although on one occasion when the author attempted this by hand, it took twenty minutes even with the assistance of the relevant manual. Finally, there were two occasions when there were insufficient channels available on the PDP-9 to enable it to attempt the connection. This condition is likely to be transient and the user could re-try in a few minutes (which, in fact, was done on these occasions).

The second group of reasons for sessions without a search were due to user intervention. The most common reason (seven times) was that the user simply gave up while the program attempted to make the connection. Since this process took a relatively short time (and we discuss the question of time later), this behaviour seems anomalous. On four occasions, the program logged in but the user was unable to supply the correct password and was therefore disconnected.

On three occasions, the user, on being advised that the program had been unable to connect on its last attempt, decided not to wait while it retried. It is likely that, depending on the reason given by the PDP-9, the user may have attempted the connection manually, but we are unable to determine that from the information recorded.

Finally, on six occasions, the PDP-9 crashed while attempting to connect. On two of these, it appears that the user re-tried after the auto-restart; on the others, one may assume that the user attempted to connect directly.

On examining the time taken for these sessions, we discover that the longest time connected to the PDP-9 was under seven minutes with a number taking about ten seconds (to be more accurate, the maximum was 350 seconds, the minimum eight seconds, the mean 135 and the standard deviation 18). The longest sessions were when the PDP-9 was able to connect to NLM but unable to log in and it is quite clear from the above discussion why this should be so. On the other hand, in cases where it is possible to obtain definite status information (even if only that the PDP-9 was unable to connect on its last attempt), the sessions were of the order of a minute. From the author's experience (and that of many other users), this is very much less than not using the PDP-9 facilities, together with the considerable advantage of a clear status message.

5.2 Searches Performed

There were ten sessions in which the user managed to perform one or more searches and we examine these in detail in this section. It must be emphasised that the statistical significance of ten sessions is low and that we are merely presenting sample data. Besides examining parameters of the searches, we are also interested, in a qualitative sense, in the ease of using the system. One way in which this may be done is to examine the number of mis-typings and mis-spellings made. Unfortunately, if local editing (i.e. on the PDP-9) were performed - as was extremely likely with the more experienced users - this is not

visible since line reconstruction is performed before the program receives the line. However, if the user utilizes the NLM editing facilities, these may be detected. In the future, this local editing will be a more common procedure with more intelligent terminals. For this reason, the result of this method of measurement is probably more representative of the future than a straight character count with central editing.

In the searches performed, only eight uses were made of the character delete facility and no use of the line delete. Occasionally (three times), an error was not picked up by the user before transmission and hence the command (e.g. "DIAGRRAM") was rejected by NLM. In all cases of deleting characters, the error was merely a single character mistyped (although in one case, the error was not detected until the end of the line and many characters had to be retyped). This is indicative of either a high degree of care or of use of the UCL editing facilities and, in fact, is probably a combination of both. One would expect the former since the searches appear to have been performed by the intermediary and not by the requester; since he is not necesarily familiar with the search terms, it is likely that considerable care will be taken.

Examining the parameters of these interactions, we find some considerable variations. These parameters are summarized in Fig. 5. Some of the meanings of the headings may not be completely clear and we explain them in detail below.

The first column refers to the number of searches performed; if the same search was performed on both the MEDLINE and CANCERLINE files, this was counted as two searches. The number of search statements (SS) refers to the actual number, excluding any commands which were given. The maximum number of search statements per search was twelve and the average was about seven.

The number of MeSH terms used averaged about one per search statement; the fairly low number is due to the number of search statements which consist of, for example, "1 AND 2", i.e. combinations of previous search statements.

As stated earlier, search statements may consist of a number of terms (either MeSH terms or the number of a previous search statement) connected together with AND, OR or AND NOT. These were used fairly often, particularly for the longer searches. In general, more connectives were used for CANCERLINE than for MEDLINE searches.

The heading "Commands" gives the number of different commands given (where, for example, "PRT", "PRINT" and "PRINT FULL" would

be considered to be the same command). In general, very few commands were used. There are twenty-three possible commands, of which the highest number used was five. The most common were, as would be expected, "PRINT", "STOP" and "FILE" (to change from MEDLINE to CANCERLINE and vice-versa). "DIAGRAM" was used fairly often as users were asked to generate a diagram of their search before printing the citations, but the other commands were used very rarely. The ones used were "SUBHEADINGS", "EXPLAIN", "NEIGHBOR", "NEWS" and "ERASEALL", the others being completely unused. EXPLODE, which is not strictly a command, was used once.

Finally, we look now at data which we can compare with the previous year (Ref. 1). These refer to such parameters as the number of lines and characters in each direction. In general, there is not much difference and such difference as there is may well be attributable to the relatively low number of interactions monitored here. The number of lines received from NLM averaged 262 per session (previously 374) and the number of characters 5220 (6344) giving an average number of characters per line of 20 (17). Similar figures for the user's interactions are 394 (1218), 34 (98) and 12 (14). The ratio of lines from NLM to lines from the user is the same as previously, namely eight.

Searches	SS	Terms	Conn	Comm	Mc	Ml	Uc	Ul	Time
1	7	4	7	2	1541	124	178	23	10:09
1	6	7	4	4	5832	249	267	34	12:24
1	1	3	0	1	1373	55	62	10	3:24
2	7	13	5	5	4177	248	241	37	15:37
1	4	3	3	2	748	74	90	13	5:49
2	15	19	15	5	6335	359	417	30	23:07
6	45	40	30	5	17476	1009	1542	98	59:28
1	4	5	1	2	1948	123	120	16	6:27
2	15	10	7	3	7339	404	250	32	19:45
2	13	21	13	5	5427	376	770	49	23:08

Figure 5: Parameters for Successful Sessions

6. Comparison With Post Office Monitoring

As we mentioned above, the BPO monitored use of the dialup lines during the one-week experiment we conducted. In this section, we discuss the results of their monitoring and, where applicable, we compare these results with those which we obtained.

The BPO monitoring consists of noting when each call is made and broken. This is then analyzed and various results tabulated. We now describe the results which are available and the values of the various parameters.

First, the entire data for one day is analyzed and the busiest hour noted and the data for this hour printed. This is done for two parameters. The first is the port usage (erlangs) and the second the calling frequency (number of calls). In the former case, there is little correlation between the results for the various days. In the earlier part of the week, the busy hour was in the afternoon (1745 to 1845 on Tuesday) while at the end, it was in the morning (1115 to 1215 on Friday). As far as the frequency of calls, the busy hour correlates well with that mentioned above on three days, but does not on two (Wednesday and Friday). As we do not have the full results, but only the busy hours, it is difficult to see why this occurred.

The same parameters (erlangs and call frequencies) are also calculated for a number of periods. These are the peak rate charge (0900 to 1300), the standard rate charge (0800 to 0900 and 1300 to 1800), the daytime (0800 to 1800), night (1800 to 0800) and finally, the 24 hour period.

The correlation between the calling rates and calling frequencies is much clearer in these periods. It is worthy of note that the values on Tuesday mornings are noticeably lower than on other mornings (by morning, we are considering the peak rate time) and this is explicable by the fact that NLM is unavailable on Tuesday mornings and hence any line usage is due to other users of ARPANET. Also, the values of the parameters on Friday night is much lower than on other nights, as might be expected, but is compensated for by an increased usage during the afternoon.

A similar pattern is noticed for the average call durations. This is very significantly lower on Tuesday morning and Friday night; this implies that the length of calls for NLM users is longer than for others.

Let us now compare these results with those we have obtained. First, it should be made clear that we are, in fact, measuring slightly different quantities and that, in order to make a

comparison in a number of cases, it is necessary to exercise judgement which may not always be valid. To give a simple example, a user could make a number of attempts to connect to a Host during one telephone call. The BPO monitoring would note this as one call whereas QUES would note it as a number of attempts to connect. The output from QUES would show this as the same user (hopefully he would not change his user name) on the same port and hence it would be reasonable to assume that it was only one telephone call. However, this is not necessarily the case and it is worth remembering that we and the BPO are, in fact, measuring slightly different data. Another problem is that we did not measure for the entire week since the computer was needed for other purposes on some occasions and hence there are times when we are unable to validate the data. Furthermore, the PDP-9 crashed on a number of occasions and there is obviously a loss of data in these cases.

However, with these provisos, there is a reasonable correlation between the data as measured by the BPO and ourselves, except that, in the case of the number of calls made, our figures are noticeably lower than those recorded by the BPO. This may well be due to calls being made (for example, wrong numbers) which were not intercepted by QUES. It is worth noting that the average call duration as measured by the BPO correlates well with the average of those shown in the table above, although with such a relatively small number of data, little significance may be placed on this.

7. Conclusions

Initially, the reason for implementing the complex system described in this Report was to overcome the problems of connecting to NLM, that is, to perform some of the functions of a network access machine. Towards the end of the project, it was decided to extend the facilities to those of a network measurement machine in order to determine those facilities which could be provided easily by such a system. Since many of the monitoring facilities would have been provided by a suitable Host system (but were not provided by NLM), it was decided not to expend much effort since the British Library had decided to implement the BLAISE system on which such monitoring could be done. However, this Report has shown some sample measurements to indicate those which can easily be performed. It must be emphasised that the data shown here are objective and should be considered in conjunction with the subjective data obtained from questionnaires and similar methods (Ref. 5).

Our conclusions may be broken down into two areas. First, there are conclusions related to the use of the NLM system and these must be considered in conjunction with the project reports (e.g. Ref. 5). Secondly, we must consider the conclusions with regard to the use of a mediating system ("Network Access Machine") such as the PDP-9.

It is clear from the results of this survey and from Ref. 5 that most searches were carried out by the intermediary who should be relatively experienced in use of the system and hence a low number of errors and a high data throughput are to be expected. This is borne out by the results.

However, perhaps the most interesting statistic, is the relative paucity of the command set used. Whether this was due to lack of knowledge of the effect of the commands or the fact that the method of use of the system did not require use of the others cannot be determined. Another significant fact was the relative simplicity of the search statements (i.e. the low use of connectives). This may well have been due to the ease of showing the search to the end-user, broken down in this way.

It is clear from comments from users that the PDP-9 greatly simplified access to NLM. This obviously would not be true in the case where the Host were accessible in a more reliable way which would enable users to access it easily. However, certain facilities provided by the PDP-9 would be of considerable use, even in this case (e.g. access to local time and connect time) and, if there were more than one computer to be accessed, the PDP-9 could perform the functions of a NAM and give the user a

unified method of access to the various machines. In addition to the ease of use, the provision of local editing (and, possibly, other local functions) should reduce the network traffic (and hence the cost).

In addition to the network access facilities, the PDP-9 provides monitoring facilities. This report has shown how the PDP-9 has been able to evaluate many aspects of the interaction. In some cases, this would duplicate monitoring performed at NLM; in others, specific parameters could be examined.

Thus the project has shown the value of providing a mediating computer for such access, allowing both access and monitoring functions. Many problems were encountered with the development of this system, both with the use of magnetic tape and the interface to ARPANET. At the end of the project, the former had been overcome and the latter significantly reduced. There were a number of developments which had been considered but not implemented (for example from knowledge of the time of day, the time of access and the place from which access ere made, the telephone cost could be calculated) and these would be of use were such a system to be implemented for a production environment.

APPENDIX I: "How to Access MEDLINE via ARPANET"

This Appendix reproduces the text of a Note which was distributed to all authorised users of the NLM system via the UCL node of ARPANET. Although it is, to a certain extent, a repetition of matter already mentioned in the main text, it was thought worthwhile to reproduce it in its entirety. The only modification has been to remove any confidential information such as the TIP telephone number.

INDRA Note No. 449

7 December 1976

Version 1.4

How To Access MEDLINE via ARPANET

by

Adrian V.Stokes

ABSTRACT

This Note gives a brief description of how to access MEDLINE via the UCL node of the ARPA Computer Network.

Dept. of Statistics and Computer Science

University College London

N.B. The information contained in this document is <u>CONFIDENTIAL</u> and is only for use by authorised ARPA/MEDLINE users

1. Introduction

This Note gives a brief description of how to access MEDLINE via the UCL node of the ARPA computer network and indicates action to be taken in the event of faults. It does not give any information on the use of MEDLINE; for this, please see Ref. 1. A "facts card" containing a summary of the information in this Note is also available (Ref. 2).

The description of the method of access is given step-by-step and it is hoped that all the common errors are noted with each step. However, for more exotic errors, there is a short section on errors in general. The first part of this Note covers the conventional method of access, namely to connect a terminal attached to the London-TIP, usually via the switched telephone network, to a pseudo-terminal on the National Bureau of Standards TIP. However, we have developed a program which runs on a PDP-9 at University College which automates the LOGIN procedure and this is also described.

2. The Communications System

The ARPA network is a heterogeneous, packet-switched, distributed computer network which has over 50 nodes, mainly in the continental USA, connected by wideband leased telephone lines (typically 50 Kbps). There are two satellite links on the network, Hawaii and Norway. Thus the user accessing MEDLINE goes via London, Norway and Washington. In the event of a line failure between London and Washington, there is, at present, the possibility of messages being re-routed via another satellite link. This goes from the London-TIP to a SIMP (Satellite IMP) at Goonhilly, Cornwall and thence to Washington. However, this link is, in fact, part of a satellite network and will only provide an alternate route to the NORSAR line for the next few months.

The UCL node of the ARPANET consists of a TIP (Terminal Interface Message Processor, a modified Honeywell 316 computer) which is connected to three "host" computers and can handle up to 63 teletype lines. It is currently connected to two PDP-9s and one PDP-11/35 computer at UCL. The first PDP-9 (Host 42) is also connected to the Rutherford Laboratory (RL) 360/195; the second will be described below (Host 106). The PDP-11 (Host 234) acts as a gateway to a satellite network. The network address of the TIP is Host 170. On the TIP, we have dial-up ports at 300/300 bps and 75/1200 bps (input/output).

For more details of the UCL node, reference is made to (3).

3·1

3. How To Access MEDLINE

(i) Dial the UCL TIP (01-xxx xxxx). You should obtain a carrier signal. If the 'phone rings, wait and you will get either a recorded message explaining why (usually the line to the USA is down; possibly because all ports are in use) or someone at UCL will answer it.

(ii) Turn the modem to "DATA" and type capital-E once or whatever other character you have been told by UCL staff (since the character depends on the terminal type; most just require the E). The TIP should reply:

LONDON TIP 371 #: 70

where the first number is the TIP system version number and the second is the port number.

If you get unintelligible characters out, it means that the TIP has made a mistake in deducing the type of terminal you are using. This can often be caused by noise on the line or by your typing a wrong initial character. If you have an acoustic coupler (e.g. a Miniport), the noise of shutting the lid is usually enough to confuse the TIP. If so, press the BREAK key and type the E again. If this does not work, hang up and try again. If there is no response at all, check your terminal - it may be in LOCAL mode !

(iii) If the LONDON-TIP monitoring service is not operating (see Section 5 if it is), type the following sequence of instructions:

@H 147	connect to NBS TIP
@R F S 1340C003	set receive socket number
@S T S 1340C002	set send socket number
@E L	set mode (echo local)
@F	Flush any characters in buffer
@P B	Open both halves of connection

The TIP should give no responses until the end when it should type:

OPEN T
OPEN B

This may sometimes be replaced by:

OPEN T B

or just

OPEN

The differences are due to the timing of the opening of the transmit and receive sides of the connection. If it either does nothing or types CLOSED, there is an error; either the socket you have chosen is already being used or it is not connected to NLM. If it does nothing for, say, half a minute, type @C and then try again with a different socket.

The "134" in the socket number (N.B. It MUST be the same number for the receive and send halves of the connection) may be replaced by 136, 140, 142 or 144. The two lower numbers operate at 10 characters per second, the other three at 30 cps and it is advisable to use one operating at the appropriate speed for your terminal if possible. If this is not possible, using a 30 cps socket with a 10 cps terminal can lead to loss of characters and other subsequent errors. However, if it is found that characters are lost using a 30 cps socket, this should be obviated using one at 10 cps.

(iv) The system at NLM, IBM's TCAM will then type:

PLEASE ENTER /LOGIN

Do as it says, type /LOGIN (N.B. including the slash), followed by a carriage return and the system will prompt you for your userid which should then be typed. The system will then say HELLO FROM EIHILL 3.

If the system persists in repeating the above message, either completely or partially, type

@E L

followed by two carriage returns, then proceed as above.

(v) At the end of the session, type:

"STOP"

(N.B. Include the inverted commas). The program will ask you to enter an @ sign for every "intellectual" search. However, this is the TIP's attention character and so, for the TIP to transmit one @, you must type a pair of them.

(vi) When TSC has logged you off, type

ac

to close the connection to NLM (the TIP should reply CLOSED or similar), then hang up your phone.

4. Errors

There are a number of faults that can occur so the procedure is not as simple as that outlined above and it is difficult to indicate the complete range of possibilities. In general, the solution to any problem is to try again, from the previous step if possible. For example, the connection may suddenly close. In this case, re-connect to the same socket and continue exactly where you left off (although you might get a spurious message saying that there are no postings for the keywords OPEN and CLOSED).

The only other relatively common problem is that the connection may go down while you are connected. This will be indicated by a

HOST NOT RESPONDING
or NET TROUBLE

message. In that case, the solution is to try again later, although you will be billed for an extra 15 minutes connect time which you did not use. However this is unlikely as our line availability is around 98 - 99%.

5. The LONDON-TIP Monitoring Service

Part of the research of the INDRA group is concerned with monitoring of ARPANET usage and this is done (in part) by means of a program running in our second PDP-9. This program is not run all the time, but is usually running in the mornings.

The program works in the following manner. When you dial up the TIP, the port to which you are attached is in a state known as "hunting", that is, it is waiting to find out the speed of your terminal. When you type the capital E, it goes into a mode which allows another user to be connected to you and this is what the PDP-9 does. In fact, it attempts to make the connection every two seconds and fails if your port is in hunting mode. When it connects to you, it asks questions (see below), then disconnects and allows you to carry on in your normal way. It attempts to reconnect to you after 20 seconds and cannot if you have either put the phone down or have connected to the host you want. It then tries to connect to you at two second intervals until it can

do so, then it repeats the questions (slightly modified).

When you first attempt to log in to the LONDON-TIP, the questions you are asked are your surname, your TIP password and the Host you wish to access. The reason for your surname is that if anything goes wrong, we know who to get in touch with. PLEASE put your real surname, although this is not checked at the moment. You will be allowed one wrong attempt at the TIP password, then you will be disconnected. For the Host, either 147 or MEDLINE will be accepted. When the questions are asked the second time (after you disconnect from NLM), you will only be asked the number of the next Host you wish to access.

To save time, when asked for your surname, you may enter your surname, TIP password and Host number, separated by 'slash' characters e.g.

SURNAME > STOKES/UCL/147

6. The MEDLINE Program at UCL

As can be seen from the above description, it is not easy to connect to NLM using the standard procedure, due to the fact that NLM is not connected to ARPANET in the usual manner. Also, in the case of any problems, it is very difficult to get accurate status information.

To overcome these problems, we have developed another program running on our PDP-9 which connects the user to MEDLINE easily, selects a suitable ID and gives specific status messages, as far as it is able. Also, the method of accessing the program is to simply state that one wishes to use MEDLINE or Host 147 to the monitoring program described above.

The program keeps up-to-date status information about MEDLINE either by remembering the status of the last user to be connected or, if there are no users, by polling the ports every five minutes. When answering to the monitoring program that one wishes to use MEDLINE, the program is called and states the last known status. If MEDLINE was available, it attempts to make the connection, trying all five ports in turn. If successful (and if it finds a port in a peculiar state, it attempts to correct this), it asks the user if he wishes to choose an ident. If not (denoted by carriage return), it will choose one that is not already in use and tell the user to which port he is connected and the ID he is using. The program is then not seen by the user, except that it modifies some of the replies from NLM, specifically deleting the prompt PROG: and replacing the prompt

USER: by the more conventional >. Also, unseen by the user, the program is performing certain monitoring functions.

A typical scenario is shown in Appendix 1.

There are a number of useful commands which may be typed at any time while the UCL MEDLINE program is running. These commands are intercepted by the PDP-9 and not sent to NLM and thus can even be sent when NLM is awaiting an answer from the terminal There are two sets of commands; the first start with the '%' character and are dealt with by the PDP-9 network software. These are used to get help information about the system and status information about the PDP-9. Currently, there are two such commands:

%H XXX,YYY	Help information
%S	Status information

In the case of the %H command, the XXX is a filename (currently the names are GATEWAY and MEDLINE; only the first three characters need be typed). If the ',YYY' is omitted, a list of the headings of the file is typed. These headings are then used as the 'YYY'. For example, if %H MED is typed, one of the headings is 'Features...' and to obtain this information, the command %H MED,FEA should be typed.

The second set of commands start with the '#' sign and are dealt with by the MEDLINE program. There are currently five of them:

#C	Print time logged in
#M	Set up initial message
#S	Print status information
#T	Print current date and time
#X	Cancel initial message

The #M and #X commands <u>must only be used by UCL staff and the Cancerline project office</u>.

7. Sending and Receiving MAIL

In a project such as this, there are significant problems communicating with many users quickly. Furthermore, if it is necessary to give technical information, the telephone is not an ideal medium. Therefore, we have set up a set of files on the ISI PDP-10 which users may interrogate often and which are updated regularly by the author. Conversely users may append messages to a similar file. To read the MAIL file, the following should be typed (after having given 86 as the Host if the monitoring program is running):

```
@C 86
login uk xxxx <space>
msg
ha              to type all headings
t3              to type message 3 (for example)
q               to exit
LOGO            log out from ISI
@c              Close the connection
```

In the above, <space> means press the space bar. All the system prompts have been removed, although, when using the system, it should be fairly obvious where to type the appropriate commands.

To send messages, a similar procedure is needed:

```
@O 86
login uk xxxx <space>
sndmsg
KIRSTEIN
AVS@@LONDON
Test message
ATTN: AVS
This is a test message
<ctrl-Z>        i.e. press CONTROL and Z
LOGO
@c
```

Obviously, receiving and sending of messages may be combined. Whenever one receives the @ prompt from ISI, one can go into the SNDMSG or the MSG subsystems. The procedure described above is extremely brief and is given in considerably more detail in the appropriate User Notice.

Appendix 1: A Typical Scenario

This Appendix shows a typical interaction with the LONDON-TIP monitoring program, including examining MAIL and sending a message. Also, this has been slightly edited to fit the page layout.

```
LONDON-TIP MONITORING SERVICE

SURNAME > STOKES
TIP PASSWORD >
UCL
HOST NUMBER > 147

WHEN I LAST LOOKED A COUPLE OF MINUTES AGO, MEDLINE WAS AVAILABLE

TERMINAL (FAST OR SLOW) >slow

PLEASE WAIT WHILE I TRY TO CONNECT YOU...

TRYING PORT #134

ID >
YOU ARE NOW CONNECTED TO PORT #134 AT MEDLINE AS USER  NLLC2
NLM PASSWORD >

HELLO FROM ELHILL3.

............

GOOD-BYE!
@c
CLOSED
OPEN R
OPEN T

LONDON-TIP MONITORING SERVICE.
 NEXT HOST NUMBER OR RING OFF NOW > 86

OK - BYE

CLOSED

@C 86
Trying...
  Open
```

```
ISI-TENEX 1.33.3, ISI-SYSTEM-A EXEC 1.53.7
@LOGIN UK XXXX
 JOB 25 ON TTY2 13-DEC-76 07:30
 PREVIOUS LOGIN: 13-DEC-76 07:10
 TENEX WILL GO DOWN THU 12-16-76 2200 TIL FRI 12-17-76 0500
 FOR PREVENTIVE MAINTENANCE
 YOU HAVE NEW MAIL
@MSG

  MSG -- Version of 1 April 1976
  Type ? for help, ? # for news

  Last read: 30-JUN-76 05:23:31; 58 msgs, 27 disk pages
  You have old messages which are 'Not Examined'


<-type all
1    7 NOV 1976      KIRSTEIN                  Medline Mesages
2   11 NOV 1976      KIRSTEIN at SRI-AI        More MEDLINE Problems

<-type 2
Mail from SRI-AI rcvd at 17-NOV-76 0438-PDT
Date: 17 NOV 1976 0440 PDT
From: KIRSTEIN at SRI-AI
Subject: More MEDLINE Problems
To:     uk at ISI
cc:     avs at LONDON

ATTN: MEDLINE

   ............Adrian

------

<-quit
Warning: Some Messages are 'Not Examined' (Confirm) Yes
Goodbye
@SNDMSG

To (? for help): KIRSTEIN@ISI

cc (? for help): avs@LONDON

Subject: Test Message

Message (? for help):
ATTN: AVS
          This is a test message
|Z
```

```
Q,S,?,carriage-return:

kirstein at ISI -- ck
avs at LONDON -- ck
@LOGO
KILLED JOB 25, USER UK, ACCT INET-UK, TTY 2, AT  12/13/76 0748
  USED 0:0:19 IN 0:18:03
@C

CLOSED
OPEN R
OPEN T

LONDON-TIP MONITORING SERVICE
NEXT HOST NUMBER OR RING OFF NOW >


(User then rang off)
```

The various programs used in the above examples are liable to change at short notice. However, all users will be notified via the MSG system of such changes and, as far as possible, they will be made self-explanatory. In case of any problems, type a question mark and you will be given whatever on-line help is available.

Appendix 2: Some Useful Telephone Numbers

01-xxx xxxx	TIP Telephone number
01-387 6768	TIP Computer Room (direct line)
01-387 7050 x 337	A V Stokes (room)
01-387 7050 x 817	" (computer room)
01-959 6665	" (home)

91

Appendix 3: Some Useful TIP Commands

There are a few extra TIP commands which may be of some use in certain circumstances. The following is a short list. In all cases, the full command is shown, but only the first character of each word need by typed.

@Protocol To Transmit	Open the transmit half of the connection
@Protocol To Receive	Open the receive half of the connection These two commands may be of some use in the case where the NBS TIP sends a message such as 'Host broke the connection Open T'. The usual @P B command is simply the combination of these two commands.
@Flush	Flush all the characters in the TIP's buffer. This performs a very crude form of editing, but is of little use since the user is not aware of which characters are still in the TIP's buffer. It may be of some use in combination with the next command.
@Transmit On Linefeed	Transmit to NLM at the end of each line. This command does not in fact do what it appears. It tells the TIP to transmit at the end of each line <u>if it can do so</u>. If it needs to transmit before, it will do so. This command cuts down the overheads in the network slightly.
@Echo Local	The usual echo state. The terminal should be in full duplex and the characters printed are, in fact, echoed from the TIP.
@Echo None	The opposite to @E L. The TIP echoing is switched off and the only way to see what is being typed is to put the terminal in HALF-DUPLEX mode. These two commands together may be useful in the case of the terminal printing corrupt characters. By using them, it is possible to determine whether the problem is in the transmit or the receive side.
@Reset	Resets the terminal to the state it was in after dialling up the TIP but before

typing the E. Therefore, the usual next step to this command is to type a capital E (or whatever the appropriate character is for your terminal).

References

1. MEDLINE Reference Manual, PB-240-126, National Library of Medicine, March 1975.

2. Stokes, A.V., "How to Access MEDLINE", ARPA Facts for UK Users, Card 4, University College London, June 1974

3. Kirstein P.T., "Status of the UCL ARPA Network Project - 1 April 1974", Indra Note 346, University College London.

4. Bates D.L., "MEDLINE - A Program to Connect Users to the MEDLINE Data Base", Indra Note 432, University College London, June 1975.

APPENDIX II: Sample Monitoring Output

```
29 OCT 76 15 06 09.1#<  1  MEDLINE PROGRAM VERSION 2.11
29 OCT 76 15 06 47.7#<  05HJL       ICE 147
29 OCT 76 15 06 47.8*<  05HJL       ICE 147
29 OCT 76 15 06 48.2B<  WHEN I LAST LOCKED A FEW MINUTES AGO,$
29 OCT 76 15 06 48.4B<  MEDLINE WAS AVAILABLE
29 OCT 76 15 06 48.8B<  TERMINAL (FAST OR SLOW) >$
29 OCT 76 15 06 51.8B>  f
29 OCT 76 15 06 51.9B<
29 OCT 76 15 06 52.0B<  PLEASE WAIT WHILE I TRY TO CONNECT YOU...
29 OCT 76 15 06 52.1B<
29 OCT 76 15 06 52.9*<  TRYING PORT #144
29 OCT 76 15 06 53.0B<  TRYING PORT #144
29 OCT 76 15 07 05.1C<  /LOGIN
29 OCT 76 15 07 08.2C>  /LOGIN
29 OCT 76 15 07 08.3C>
29 OCT 76 15 07 09.3C>
```

ACRONYMS

ARPA	Advanced Research Projects Agency
BL	British Library
BLL	British Library, Lending Division
BLRD	British Library, Research and Development Dept.
bps	Bits per second
CADC	Computer Aided Design Centre (Cambridge)
HEP	High Energy Physics
IMP	Interface Message Processor
INDRA	InterNetwork Display and Remote Access
ISI	Information Sciences Institute
MEDLINE	MEDLARS On-Line
NAM	Network Access Machine
NBS	National Bureau of Standards
NLM	National Library of Medicine
NMM	Network Measurement Machine
PSTN	Public Switched Telephone Network
RL	Rutherford Laboratory
RSRE	Royal Signals and Radar Establishment
SIMP	Satellite Interface Message Processor
STEIN	Short-Term Experimental Information Network
TIP	Terminal Interface Message Processor
UCL	University College London
UCLA	University of California at Los Angeles

References

1. Stokes, A.V., "The INDRA Network Measurement Project", INDRA TR-30, Department of Statistics and Computer Science, University College London, 1977.

2. Holmes, P.L., "An Approach to the Development of Library and Information Networks with Special Reference to the U.K.", AGARD CP-158, March 1975.

3. Holmes, P.L., "On-Line Information Retrieval - An Introduction and Guide to the British Library's Short Term Experimental Information Network Project", British Library Research and Development Department, October 1976.

4. Kenney, S.B., "Description and Analysis of the Use of the UCL ARPANET Node During 1976", INDRA Note 602, Department of Statistics and Computer Science, University College London, 1977.

5. Pollitt, A.S., "CANCERLINE Evaluation Project - Final Report", University of Leeds School of Medicine Medical Library, May 1977.

6. Stokes, A.V., "A User's Guide to Babbage", INDRA TR-13, Department of Statistics and Computer Science, University College London, 1977.

7. Stokes, A.V., "Specification of a Routine to Connect Users to MEDLINE", INDRA Note 402, Department of Statistics and Computer Science, University College London, 1975.

8. Bates, D.L., "MEDLINE - A Program to Connect Users to the MEDLINE Data Base", INDRA Note 432, Department of Statistics and Computer Science, University College London, 1975.

9. --, "On-Line Services Reference Manual", PB-240-126, National Library of Medicine, 1975.

10. --, "Medical Subject Headings - Annotated Alphabetic List 1975", PB-240-829, National Library of Medicine, 1975.

11. Stokes, A.V., "XESTATS - A User Guide (revised)", INDRA Note 639, Department of Statistics and Computer Science, University College London, 1977.

12. Stokes, A.V., "The Babbage System Library", INDRA TR-35, Department of Statistics and Computer Science, University College London, 1977.

Acknowledgements

We wish to acknowledge support by the British Library under grant SI/G/093. Also, such research would not be feasible without support of the many bodies for the ARPA project at University College, in particular, the Science Research Council (B/RG/5981 and 67022), the US Advanced Research Projects Agency (N00014-74-C-0240 and N00014-77-B-0005) and the UK Ministry of Defence (AT/2047/064).

We also wish to acknowledge the contributions of many people of the INDRA group at University College to this project, in particular, to David Bates and Roger Gould who did much of the coding and debugging of the MEDLINE and QUES programs and Howard Lethbridge who wrote the analysis programs for the MEDLINE data as part of his M.Sc. project. Some examples and Fig. 4 of this report are extracted from his thesis.